基于表情符注意力机制的微博情感分析模型 *

谭　皓 [1]，邓树文 [2]，钱　涛 [2]，姬东鸿 [1]

(1. 武汉大学 计算机学院, 武汉 430072; 2. 湖北科技学院 计算机学院, 湖北 咸宁 437100)

摘　要：为了对中文微博进行有效的情感极性识别，基于表情符能改变或加强微博文本的情感极性这一认知事实，提出基于表情符注意力机制的微博情感分析神经网络模型。该模型在使用双向循环神经网络模型（Bi-LSTM）学习文本的特征表示时，利用表情符注意力机制，得到文本结合表情符后新的特征表示，从而实现微博情感识别。实验结果显示，与输入纯文本和表情符的 Bi-LSTM 模型相比，基于表情符注意力机制的模型准确率提高了 4.06%；与仅输入纯文本的 Bi-LSTM 模型相比，基于表情符注意力机制的模型准确率提高了 6.35%。

关键词：表情符；微博；情感分析；注意力机制

中图分类号：TP391　　doi: 10.3969/j.issn.1001-3695.2018.03.0152

Emoji-attentional neural network for microblog sentiment analysis

Tan Hao[1], Deng Shuwen[2], Qian Tao[2], Ji Donghong[1]

(1. *School of Computer*, *Wuhan University*, *Wuhan* 430072, *China*; 2. *School of Computer Science & Technology Hubei University of Science & Technology*, *Xianning Hubei* 437100, *China*)

Abstract: In order to effectively identify the sentiment polarity of Chinese microblogs, based on the cognitive fact that emojis can change or enhance the sentiment polarity of the texts, this paper proposed an emoji attentional neural network model for microblog sentiment analysis. This model firstly use a Bi-LSTM model to learn the feature representation of the text, and then an emoji-based attention mechanism is implemented to obtain a new feature representation after combining text with the emojis, so as to recognize the sentiment polarity of the microblogs. The experiments show that compared with the Bi-LSTM model that inputs plain text and emojis, the accuracy rate of the emoji-attentional model is increased by 4.06%; compared with the Bi-LSTM model that only inputs plain text, the accuracy rate of emoji-attentional model is increased by 6.35%.

Key words: emoji; microblog; sentiment analysis; attention mechanism

0　引言

现如今，微博是人们交流信息、表达情感和观点最重要的媒介之一，已有大量的研究工作聚焦于微博文本，如舆情分析[1]，突发事件检测[2]，观点挖掘[3]和情感分析[4,5]等，其中情感分析是微博研究的基础及重点话题之一。当前情感分析采用的主要方法包括支持向量机[6]，决策树[7]，及神经网络[8-10]等模型。为表达情感和观点，人们在微博中使用了大量的表情符，因此在情感分析的任务中，许多研究者也把表情符作为重要的因素融入到模型中。常用的方法是把表情符作为模型特征[11-12]或一种自然标注[13-14]，用于扩充训练语料。然而研究[13]显示表情符直接当作情感标签会产生很多噪声。

社会及认知学领域研究认为表情符类似于人类交流中的非言语成份，如面部表情，姿态等，表达交流者之间的情感[15]，他们认为：在社交媒体中，表情符对文本的情感表达产生最重要影响，甚至能改变文本的情感极性。如下面两个例子:

①明天要上学😭

②妈妈给我的生日礼物😭

文本的情感变化如图 1 所示。

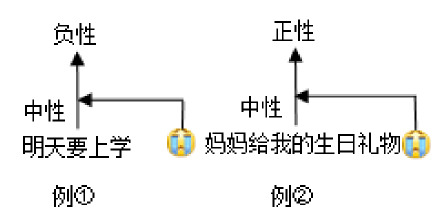


图 1　微博文本的情感变化

收稿日期：2018-03-05；**修回日期：**2018-04-12　**基金项目：**国家自然科学基金面上项目（61772378）；国家社会科学基金重大项目（11&ZD189）；湖北省自然科学基金一般面上项目（2018CFB690）

作者简介：谭皓（1993-），男，四川广安人，硕士研究生，主要研究方向为自然语言处理；邓树文（1979-），男（通信作者），湖北咸宁人，讲师，硕士，主要研究方向为情感分析、机器学习（rolex_deng@163.com）；钱涛（1975-），男，副教授，博士，主要研究方向为自然语言处理、机器学习；姬东鸿（1967-），男，教授，博导，博士，主要研究方向为自然语言处理、数据挖掘等.

可以看出，表情符对文本的情感表达起着重要作用。但微博的情感表达并非文本与表情符的简单相加，例如两个例子虽然具有相同的表情符，但在不同的文本下，其产生的作用是不同的。这种作用类似于神经网络的注意力机制[16]，在一个微博句子中，每个词语的信息量权重是不同的，表情符同时也对文本中词语的权重产生影响，改变文本的表达，进而改变文本的情感极性表达。

因此本文结合这两者，提出基于表情符注意力机制的神经网络模型。图 2 给出了本文所提模型框架例子。把微博分成两类信息：文本和表情符。模型首先对输入文本采用双向循环神经网络网络模型(Bi-LSTM)学习出文本的低维特征表示，然后结合表情符的向量，输入表情符注意力网络层，改变文本中单词的权重，进而得到文本的特征表示，最后利用分类方法得到文本的情感标签。

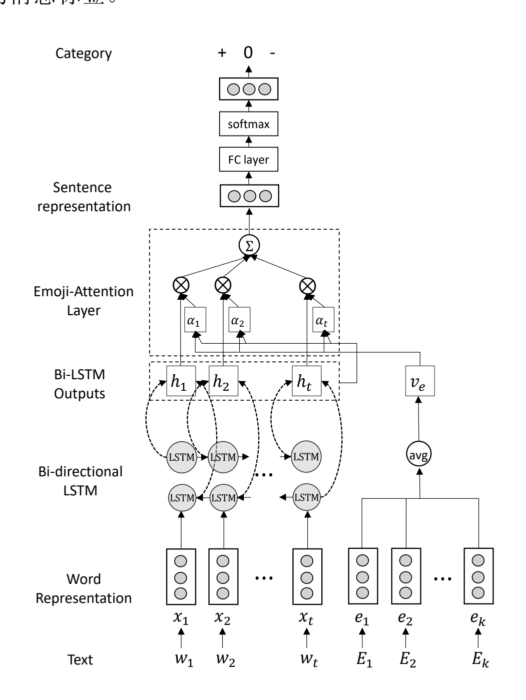


图 2　基于表情符注意力机制的模型框架

本文的主要贡献如下：a）构建了一个基于表情符的微博情感标注语料。该语料共包含 6905 条微博;b）不同于以往仅将表情符作为单个特征的工作，本文在处理语料时，将微博中的表情符与词语同时训练，得出含有上下文信息的表情符词向量，并提出基于表情符注意力机制的神经网络模型。实验结果也显示了该模型的有效性。

1　相关工作

情感分析是对带有情感色彩的主观性文本进行分析、处理、归纳和推理的过程。例如对电影的评论，产品的评价，情感分析有助于客观评价产品。其通常被看作是一个分类问题，各种分类模型被广泛应用该类任务，如 SVM、决策树、神经网络等，常用的特征包括单词、单词数、词性、句法等，以及人工标注的情感词典。

近年来神经网络模型飞速发展，由于避免了复杂的特征工程，已广泛应用于情感分析。大多数方法采用 RNN,CNN 等模型。Chen 等人[16]提出了一个结合用户-产品的注意力机制，用以完成情感分析任务。Zhang 等人[17]提出了门机制。这些模型主要把文本作为输入，没有考虑表情符对文本的作用。

在社交媒体如新浪微博、twitter 中，存在着大量的表情符号，用于表达用户的情感及观点。当前对表情符的使用策略主要包括三种：

第一类策略[13,14]是将表情符看做是一种自然标注，该策略假设表情符独立的表达了用户的情感和观点。Purver 等人[18]使用 Twitter 数据中的标签（hashtags）和表情（emotions）来产生训练数据集。实验结果表明，本方法在快乐（hapiness）、伤心（sadness）、愤怒（anger）三种表情作为情绪标签时比较有效，而对于其他类型的表情效果不佳。Davidov 等人的研究[13]显示，由于表情符存在歧义，即同一个表情符也可是正性，也可能是负性，因此利用表情符建立的训练语料包含许多噪音，可能会训练模型产生副作用。

第二类策略[11, 12]是将表情符作为一种特征融入到分析模型中。Jiang 等人[19]在 SVM 模型中融合了微博结构特征、句子结构特征与表情特征，对微博进行正面、中性、负面的三类情感分类。这类策略也没有体现表情符对文本的情感作用。

第三类策略是把表情符和文本看作两个并列的信息源。Hogenboo 等人[20]将社交媒体文本分成表情符和文本两部分，然后分别采用不同的模型计算出各自的情感，最后两个情感线性组合得到最终的文本情感。

上述三种策略，虽然在模型中考虑了表情符，但没有考虑到表情符对文本的作用机制。

注意力机制早期运用于图像处理[21]，近来也被逐步运用于自然语言处理，其作用是从众多信息中选择出关键信息。在自然语言处理中，研究显示注意力机制对机器翻译[22]、问答系统[23]、情感分析[16]性能都有较大提升。本文提出基于表情符注意力机制的神经网络模型，较好地模拟了表情符对文本的情感作用这一认知事实。

2　模型

2.1　基于 Bi-LSTM 模型的情感分类模型

双向长短时记忆模型（Bi-LSTM）是一种特殊循环神经网络（RNN）模型，主要目的用来处理序列数据，其在自然语言处理中取得了广泛应用。在情感分类任务中，该模型通常被运于学习句子的表示，然后依据表示对文本作情感分类。Yang 等人[24]使用了 Bi-LSTM 完成文档分类，取得了较好的效果。下面简要介绍 Bi-LSTM 模型。

LSTM 是特殊的循环神经网络,能够提升神经网络接受输入信息及训练数据的记忆机制,让输出结果得到大幅度的提升。

对于短文本中的第 t 个单词 w_t ,首先模型将 w_t 映射到一个词向

量 $x_t \in \mathbb{R}^d$，一个 LSTM 单元的输入为:单词 x_t，上一输出隐状态 h_{t-1}，上一内存状态 c_{t-1}，输出为 h_t，c_t。具体公式如下:

$$\begin{aligned} i_t &= \sigma(W_i x_t + U_i h_{t-1} + b_i) \\ f_t &= \sigma(W_f x_t + U_f h_{t-1} + b_f) \\ \tilde{c}_t &= \tanh(W_c x_t + U_c h_{t-1} + b_c) \\ c_t &= f_t \odot c_{t-1} + i_t \odot \tilde{c}_t \\ o_t &= \sigma(W_o x_t + U_o h_{t-1} + b_o) \\ h_t &= o_t \odot \tanh(c_t) \end{aligned} \tag{1}$$

对于普通的 LSTM，其存在一个缺点，即只能正向读取文本，因此本文使用能够双向读取文本的 Bi-LSTM 模型。Bi-LSTM 包含一个正向的 $\overrightarrow{\text{LSTM}}$，由 x_1 读取到 x_T，以及一个反向的 $\overleftarrow{\text{LSTM}}$，由 x_T 读取到 x_1：

$$\begin{aligned} x_t &= W_e w_t, t \in [1,T] \\ \overrightarrow{h_t} &= \overrightarrow{\text{LSTM}}(x_t), t \in [1,T] \\ \overleftarrow{h_t} &= \overleftarrow{\text{LSTM}}(x_t), t \in [T,1] \end{aligned}$$

经过这一层，模型将一个单词 x_t 转换为一个隐变量，将 $\overrightarrow{h_t}$ 与 $\overleftarrow{h_t}$ 拼接起来即为单个单词的表示，即 $h_t = \overrightarrow{h_t} \oplus \overleftarrow{h_t}$。由此得到了 $[h_1, h_2, ..., h_T]$，再将其传入一个平均池化层，则得到单个句子的表示 s。

2.2 基于表情符注意力机制的 Bi-LSTM 情感分类模型

为了表示表情符对文本的作用机制，本文提出表情符注意力机制，以建模表情符对文本的作用。对于一条微博，其每个词语对于情感极性影响是不同的，与表情符作用的效果也是不同的。表情符注意力机制，用以衡量与表情结合后，微博中的词语的重要程度权重。

对一条微博文本：$\{w_1, w_1, ...w_T; E_1, E_2, ..., E_k\}$，其中 w_i 为微博中经过分词后的词语，E_j 表示微博中的表情符。首先，w_i 与 E_j 都转换为词向量的表示 $x_i \in \mathbb{R}^d, e_j \in \mathbb{R}^d$。这里 d 为词向量的维数。表情符的向量表示为

$$v_e = \frac{1}{k}\sum_{t=1}^{k} e_t \tag{2}$$

由于许多微博用户都会在同一条微博中发表多个同样的表情，此处对微博中的表情符向量取平均值，以防止单个表情符的权重过大，影响模型的效果。

不同于 3.1 节，在本节中，令 $[h_1, h_2, ..., h_T]$ 为经过 Bi-LSTM 神经网络后，微博中纯文本 $\{w_1, w_1, ...w_T\}$ 的表示。则整个句子的表示如下：

$$s = \sum_{t=1}^{T} \alpha_t h_t \tag{3}$$

此处 α_t 表示了句子第 t 个词结合表情符后的重要程度，即模型表示的第 t 个词在句子中的注意力权重，α_t 的定义如下：

$$\alpha_t = \frac{\exp(\text{score}(h_t, v_e))}{\sum_{j=1}^{T} \exp(\text{score}(h_j, v_e))} \tag{4}$$

score 为衡量单词与表情符组合后的重要程度的打分函数，score 的定义如下：

$$\text{score}(h_t, v_e) = \mathbf{v}^{\text{T}} \tanh(W_H h_t + W_E v_e + b) \tag{5}$$

其中：W_H, W_e 为权重矩阵，b 为偏置向量，$\mathbf{v}$ 为权重向量，$\mathbf{v}^{\text{T}}$ 表示 $\mathbf{v}$ 的转置。

2.3 优化目标

在引入表情符注意力作用机制后，得到文本的最终向量表示 s。为了对文本作情感分类，模型使用一个非线性函数把文本表示映射到一个目标类别空间。

$$d_c = \tanh(W_c s + b_c) \tag{6}$$

然后，采用 softmax 函数来获取其情感分布。

$$p_c = \frac{\exp(d_c)}{\sum_{k=1}^{C} \exp(d_k)} \tag{7}$$

其中：C 表示情感标签个数，p_c 表示情感标签 c 的预测概率。

模型采用交叉熵损失作用模型的损失函数。若 D 表示训练微博集，则基于表情符注意力机制的模型损失函数表示如下：

$$L = -\sum_{d \in D}\sum_{c=1}^{C} p_c^g(d)\log(p_c(d)) \tag{8}$$

3 实验

3.1 数据集

为了训练和测试语料，从新浪微博抽取了 2017 年 4 月份包含有表情符的 10 万条微博，使用 jieba 中文分词器进行分

词，替换掉微博中的网址、用户及话题标签等。过滤掉长度小于 5 的微博文本。然后随机抽取其中的 1 万多条微博文本作为待标注微博，同时要求每个表情符出现的次数大于 10 次，去除重复以及表情符过少的微博后得到 6905 条语料。在标注微博时，其情感标签分为三类：正性、中性和负性。语料及标注统计如表 1 所示。

表 1 语料统计

语料	正性	中性	负性
6905	3113	1926	1866
不同表情符个数	表情符出现总次数	个数/实例	
109	14898	2.15	

3.2 实验设置

为了得到微博中词语与表情符的向量表示，本文采用抽取的 100 M 微博语料进行预训练。对上述语料经过如 4.1 节步骤的分词与预处理后，普通词语与表情符号作为同类单词看待。单词向量表示采用 SkipGram 模型[25]预训练，词向量表示维度为 300。由此便可得到微博词语的向量表示与包含上下文信息的表情符向量表示。若模型运行时出现不在词向量词表中的词语或表情符，采用随机初始化。

微博文本的表示采用 Bi-LSTM 来学习。每个隐结点的向量为 100 维，微博每个单词的表示由双向的隐向量连接而成，因此每个单词的输出表示为 200 维。训练时采用 Adadelta 方法[26]对参数进行优化。在每一次优化过程中，模型将更新权重矩阵、偏置向量、单词向量、表情符向量。训练次数为 100 次。

实验时标注语料按 8:1:1 分别分为训练集、开发集和测试集。采用准确度(accuracy)来表示评估实验的性能。

3.3 实验结果

本实验的主要目标是测试表情符对文本的作用，因此在实验中给出三个基线。

a）Emoji-based,仅用表情符自身极性作为判断依据。每个表情符的极性由人工标注得出。

b）Bi-LSTM-text,纯文本输入 Bi-LSTM 情感分析网络模型。去掉微博中的表情符，以纯文本作为网络的输入，Bi-LSTM 得到的微博表示直接作为判断模型的特征向量。优化目标及训练方法与本文模型相同。

c）Bi-LSTM-text-emoji,纯文本+表情符输入的 Bi-LSTM 情感分析网络模型。保留微博中的表情符，把表情符和文本同时作为网络的输入，Bi-LSTM 得到的微博表示直接作为判断模型的特征向量。优化目标及训练方法与本文模型相同。

3.4 实验分析

表 2 显示了各个模型的识别准确度。从表中可以发现，把表情符引入模型后，识别性能相比只输入纯文本的模型 68. 51%，提升至 70. 80%。主要是因为在微博文本，存在许多搞怪、自嘲、及讽刺的文本，表情符的加入，引起文本的情感极性的改变，但单纯从文本识别不出来。比较基于注意力模型与非注意力的表情符模型，基于注意力的模型准确率上升至 74. 86%。可以看出基于表情符的注意力机制能更为有效地表示文本的情感。

在表 2 中，中性情感的微博准确率依然较低，这是因为中性微博中含有部分具有倾向的词语和表情符，同时文本本身又不具有倾向性；另一个原因是部分用户对于表情符的使用比较随意，使用与正文无关的表情符。以下两句话是实际上为中性，但识别为正面的微博：

人生就像一座火山，当你生气和愤怒时它会爆发，当你高兴和欢乐时它😍会喷出地下泉，浇灌脚下的花花草草。

🍻不敢大意…

表 3 显示基于表情符注意力机制模型要好于表情符作特征的模型。为了验证注意力模型对不同情感极性微博的影响，表 3 给出对不同情感极性的微博的识别精度。可以看出添加表情符后，极性为负的准确率有较大提升，同时也超过了极性为正的识别率，而中性极性准确率最低，说明负性表情符对文本极性影响最大；表情符对中性微博准确率影响不大，但显著提升了召回值。

下面给出几个例子，Emoji-Attention

Bi-LSTM（LSTM-EA）模型能够识别，而 Bi-LSTM-text+emoji（LSTM-E）没有识别出来。

这里 1)与 2)都是明显的反讽的例子，此时 LSTM-E 模型仅将各词语的向量平均起来作为分类器的输入，产生了错误的预测。而 LSTM-EA 模型由于考虑到各词语的权重不同，预测正确。而例子 3)与 4) LSTM-E 预测错误也是因为只考虑了微博中一些负面词语，而忽视了整体情感需要整体的作用。

由于在微博中存在许多反讽以及类似上面“话锋一转”的例子，因此本文所提出的模型是具有优越性与实用价值的。

上面实验说明了本文所提模型结果性能显著优于仅考虑纯文本的模型，也优于只将表情符作为特征的模型，其主要原因是表情符改变或加强了文本的极性，同时表情符注意力模型也发掘出了文本内部的语义权重，提升了识别效果。

表 2 情感识别结果

系统	Accu(%)
表情符本身极性	56.49
Bi-LSTM-text	68.51
Bi-LSTM-text-emoji	70.80
Emoji-Attention Bi-LSTM	74.86

表 3 表情符注意力机制模型的不同情感文本的识别结果。

系统	类别	P（%）	R（%）	F（%）
Bi-LSTM-text	正性	76.05	71.11	73.49
	中性	66.57	51.71	58.21
	负性	60.64	80.00	68.99
Bi-LSTM-text +emoji	正性	73.10	77.44	75.21
	中性	64.46	64.06	64.26
	负性	74.75	69.57	72.07
Emoji-	正性	78.21	76.57	**77.38**

Attention	中性	65.35	71.66	**69.36**	Bi-LSTM	负性	80.82	74.49	**77.53**

表 4　部分模型识别不同的结果

	文本	LSTM-EA	LSTM-E	Truth
1)	我不介意秀恩爱，但我介意秀恩爱刷屏，对不起单身狗就是这么直接靴靴	负面	正面	负面
2)	不好意思，我悟性低，你的当头一棒并没有敲醒我，反而敲走了我对你的敬爱和感激，谢谢	负面	正面	负面
3)	【两车当街对撞互不相让一车主冲下车后画风突变！】3 日，网友晒出视频：河南驻马店某路口，一红一白两车当街顶牛，互不相让。在两车连续互撞两次后，红车司机下车与白车司机交谈。随后画风突变，红车司机上车，白车倒车，两人面带笑容，各自离开！PS：驻马店的老铁，没法理解~...	正面	负面	正面
4)	哈哈哈～想起上次南坪大雨 3 天出不来坐阳台上钓鱼的，我大 cq 人民真可爱顺带鄙视某些想吃人血馒头胡乱放飞的//@暮冥战机:感谢旅客们的支持和理解！	正面	负面	正面

4　结束语

本文以表情符对文本的情感表达起作着重要作用这一认知事实，提出基于表情符注意力机制的情感分析模型。在模型中，一个句子中的每个词语结合表情符后，将会被施以不同的权重对待，由此使得一个句子的语义能得到更精确的表达。实验结果显示，表情符注意力机制能有效识别微博情感极性，并且准确率与 F 值都超越了仅将表情符作为特征加入神经网络的方法，取得了较好的结果。

本文仍有一些不足，在标注时，对微博的极性标注只有三类，类别较少。另外部分用户的微博文本与表情使用较为随意，也影响了模型的表现。下一步将研究基于表情符的细粒度的微博情感分析。

参考文献：

[1] Kannangara S. Mining Twitter for fine-grained political opinion polarity classification, ideology detection and sarcasm detection [C]// Proc of the 11th ACM International Conference on Web Search and Data Mining. New York: ACM Press, 2018: 751–752.

[2] Shi Si, Jin Dawei, Goh T T. Real-time public mood tracking of Chinese microblog streams with complex event processing [J]. IEEE Access, 2017, 5: 421–431.

[3] Yin Fulian, Zhang Beibei, Huang Bochen, *et al.* An public opinion trend prediction method based on neural network algorithm [C]// Proc of the 2nd IEEE International Conference on Computational Intelligence and Applications. Piscataway, NJ: IEEE Press, 2017: 440–444.

[4] Peng Haiyun, Cambria E, Hussain A. A review of sentiment analysis research in Chinese language [J]. Cognitive Computation, 2017, 9 (4): 423–435.

[5] Cortis K, Freitas A, Daudert T, *et al.* Semeval-2017 task 5: fine-grained sentiment analysis on financial microblogs and news [C]// Proc of the 11th International Workshop on Semantic Evaluation. Stroudsburg, PA: ACL, 2017: 519–535.

[6] Pang Bo, Lee L, Vaithyanathan S. Thumbs up?: Sentiment classification using machine learning techniques [C]// Proc of ACL Conference on Empirical Methods in Natural Language Processing. Stroudsburg, PA: ACL, 2002: 76-86.

[7] Prasad S S, Kumar J, Prabhakar D K, *et al.* Sentiment classification: an approach for indian language tweets using decision tree [C]// Proc of International Conference on Mining Intelligence and Knowledge Exploration. Cham: Springer, 2015: 656-663.

[8] Tang Duyu, Qin Bing, Liu Ting. Document modeling with gated recurrent neural network for sentiment classification [C]// Proc of Conference on Empirical Methods in Natural Language Processing. Stroudsburg, PA: ACL, 2015: 1422-1432.

[9] Dos Santos C N, Gatti M. Deep convolutional neural networks for sentiment analysis of short texts [C]// Proc of the 25th International Conference on Computational Linguistics: Technical Papers. Stroudsburg, PA: ACL, 2014: 69-78.

[10] Tai Kaisheng, Socher R, Manning C D. Improved semantic representations from tree-structured long short-term memory networks [EB/OL]. (2015) [2018-04-10]. https://arxiv. org/pdf/1503. 00075.

[11] Zhao Jichang, Dong Li, Wu Junjie, *et al.* Moodlens: an emoticon-based sentiment analysis system for chinese tweets [C]// Proc of the 18th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. New York: ACM Press, 2012: 1528-1531.

[12] Ptáček T, Habernal I, Hong Jun. Sarcasm detection on czech and english twitter [C]// Proc of the 25th International Conference on Computational Linguistics: Technical Papers. Stroudsburg, PA: ACL, 2014: 213-223.

[13] Davidov D, Tsur O, Rappoport A. Enhanced sentiment learning using twitter hashtags and smileys [C]// Proc of the 23rd International Conference on Computational Linguistics: Posters. Stroudsburg, PA: ACL, 2010: 241-249.

[14] Go A, Bhayani R, Huang Lei. Twitter sentiment classification using distant supervision, CS224N [R]. Stanford, 2009.

[15] Walther J B, D'Addario K P. The impacts of emoticons on message interpretation in computer-mediated communication [J]. Social Science

Computer Review, 2001, 19 (3): 324-347.

[16] Chen Huiming, Sun Maosong, Tu Cunchao, *et al.* Neural sentiment classification with user and product attention [C]// Proc of Conference on Empirical Methods in Natural Language Processing. Stroudsburg, PA: ACL, 2016: 1650-1659.

[17] Zhang Meishan, Zhang Yue, Vo D T. Gated neural networks for targeted sentiment analysis [C]// Menlo Park, CA: AAAI Press, 2016: 3087-3093.

[18] Purver M, Battersby S. Experimenting with distant supervision for emotion classification [C]// Proc of the 13th Conference of the European Chapter of the Association for Computational Linguistics. Stroudsburg, PA: ACL, 2012: 482-491.

[19] Jiang Fei, Cui Anqi, Liu Yiqun, *et al.* Every Term Has Sentiment: Learning from emoticon evidences for Chinese microblog sentiment analysis [C]// Natural Language Processing and Chinese Computing. Berlin: Springer, 2013: 224–235.

[20] Hogenboom A, Bal D, Frasincar F, *et al.* Exploiting emoticons in sentiment analysis [C]// Proc of the 28th Annual ACM Symposium on Applied Computing. New York: ACM Press, 2013: 703-710.

[21] Mnih V, Heess N, Graves A. Recurrent models of visual attention [C]// Advances in Neural Information Processing Systems. 2014: 2204–2212.

[22] Bahdanau D, Cho K, Bengio Y. Neural machine translation by jointly learning to align and translate [EB/OL]. (2016) [2018-04-10]. https://arxiv. org/pdf/1409. 0473.

[23] Hermann K M, Kocisky T, Grefenstette E, *et al.* Teaching machines to read and comprehend [C]// Advances in Neural Information Processing Systems. 2015: 1693-1701.

[24] Yang Zichao, Yang Diyi, Dyer C, *et al.* Hierarchical attention networks for document classification [C]// Proc of Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Stroudsburg, PA: ACL, 2016: 1480-1489.

[25] Mikolov T, Sutskever I, Chen Kai, *et al.* Distributed representations of words and phrases and their compositionality [C]// Advances in Neural Information Processing Systems. 2013: 3111-3119.

[26] Zeiler M D. Adadelta: an adaptive learning rate method [EB/OL]. (2012-12-22) [2018-04-10]. https://arxiv. org/pdf/1212. 5701.